LARS Print 080172

Data Handling and Analysis for the 1971 Corn Blight Watch Experiment

by
P.E. Anuta, T.L. Phillips and
D.A. Landgrebe

Presented at the National Telecommunications Conference, Houston, Texas, December, 1972.

The Laboratory for Applications of Remote Sensing

Purdue University
Lafayette, Indiana

DATA HANDLING AND ANALYSIS FOR THE 1971 CORN BLIGHT WATCH EXPERIMENT

BY

Paul E. Anuta

Terry L. Phillips

David A. Landgrebe

Laboratory for Applications of Remote Sensing

Purdue University

Lafayette, Indiana

INTRODUCTION

In the summer of 1971 the Laboratory for Applications of Remote Sensing at Purdue University conducted a pioneering experiment in the application of remote sensing technology to a large scale earth resource surveilance problem. In 1970 a plant disease called the Southern Corn Leaf Blight infected large areas of corn production in the United States and it is estimated that total national corn production was reduced by 15% due to this disease.[1] In the winter of 1970, an experiment was designed in cooperation with the U. S. Department of Agriculture, NASA, and the University of Michigan Institute of Science and Technology's Willow Run Laboratories for the summer of 1971 to study the feasibility of detecting and mapping this disease using remote sensing techniques. In this paper, the Corn Blight Watch Experiment will be described from the point of view of data handling, data analysis and interpretation procedures used.

Planning activity early in 1971 resulted in the concept of a central data reduction center where all data preprocessing and processing functions would be performed. Three data sources were defined for the experiment. The first was the ground data collection activity to be carried out by county agents under the auspices of the U. S. Department of Agriculture. The second was aerial photography to be gathered by the NASA RB57f high altitude (50,000 ft.) surveilance aircraft. The third input was airborne multispectral scanner data from the University of Michigan C-47 aircraft operating at an altitude of 5000 feet. The advantage of a central site was that maximum communication between the individuals performing seperate data reduction functions would be possible. Due to limitation on resources at LARS modification of the central data center concept was required. As it turned out several centers of data processing activity existed for the experiment. Ground observations of selected test sites throughout a seven state region in the U. S. Corn Belt were coordinated through county ASCS (Agricultural Stabilization and Conservation Service) and CES (Cooperative Extension Serv.) offices. The data was sent to the USDA Statistical Reporting Service (SRS) at the state level and then on to the Washington, D. C. SRS office where all editing, collating and preliminary blight severity analysis was performed. The results were then sent to LARS at Purdue. At LARS the other 15 flightlines were analysed along with 210 segments of high altitude photography. LARS was responsible for collating and analyzing all interpreted results and for reporting them to SRS in Washington and to other participants in the experiment. Table 1 contains a list of the data centers involved in the experiment.

In this paper the overall Corn Blight Watch Experiment data flow is described and the organization of the LARS/Purdue data center is discussed. Data analysis techniques are discussed in general and the use of statistical multispectral pattern recognition methods for automatic computer analysis of aircraft scanner data is described. Some of the results obtained are discussed and the implications of the experiment on future data communication requirements for earth

resource survey systems is discussed in conclusion.

Data Flow

The corn blight watch experiment (hereafter referred to as CBW) was the first time a large scale quasi-operational remote sensing experiment had been conducted by LARS. Thus, we were faced with the problem of establishing data handling procedures to control the huge data flow expected. Therefore, a data storage and retrieval system was established and maintained to aid in accessing the collected data. The purpose of the system was to:

- Store data products in an organized library.
- Maintain a record of all data stored and reduced for future access.
- Report to the photo interpreters and multispectral data reduction teams all required information in a format that allowed the simplest access possible.
- Record data reduction results from photographic and multispectral scanner data reduction teams and merge the results with the data collected.
- Report data reduction results to the Statistical Reporting Service in Washington and to other participants in the experiment.

The Corn Blight Watch Experiment was conducted in three phases. The data flow plan is best presented by describing the transfer of data from data acquisition centers to and between the data processing centers for each phase. The data flow plan is graphically shown on diagrams which include each center and each transfer of data between the centers. Each center is shown on the diagrams as a node with an abbreviation for the center. These abbreviations are identified in Table 1.

Phase 1

During Phase 1, baseline data for the entire Corn Blight Watch Experiment were collected between April 15 and May 15. Data Acquisition for Phase 1 included the collection of black and white photography by ASCS enumerators. Processing included the reduction of photography to a scale of 1:20,000, the outlining of tracts and fields on the reduced photography, and the reporting of farm operator interviews through SRS to the data reduction center at LARS. Figure 1 shows the data flow for Phase 1.

NASA flew two six-inch focal length cameras at 50,000 feet to obtain two original exposures of 36 flightlines containing the 210 samples sites. The scale of this black and white panchromatic photography was 1:100,000. The segments were located, enlarged to a scale of 1:20,000, and three identical prints of each site were delivered to SRS in Washington. A duplicate set of the original photography was also sent to ASCS. All baseline photography was collected with less than 10% cloud cover present. For some segments not photographed within the time period or cloud cover conditions, existing ASCS photography was used.

At SRS the segment was outlined on one copy of each print and this was sent with initial interview forms to the Agricultural Stabilization and Conservation Service county enumerators. Each farm operation in the segment was outlined and visited by the ASCS enumerator. During the interviews a field ID was assigned to each field, field boundaries and ID annotations were added to the photograph, and the initial interview forms were filled out for each field in the segment. The annotated photographs and completed initial interview forms were delivered to SRS. The annotations were copied onto the other two sets of prints and the data on the forms were coded, punched, edited, and recorded in digital format. One set of baseline photographs and a digital copy of the initial interview data were sent to the Data Reduction Center at LARS/Purdue. A second set of baseline photographs was sent to the Cooperative Extension Service in each state where segments were located.

Phase 2

In Phase 2, between May 10 and May 30, color IR photography was collected over the 210 segments and multispectral scanner data were collected over the 30 intensive study area segments. This data was analyzed for soils characteristics to provide soils background information for corn fields in the segments. The flow diagram for Phase 2 is shown in Figure 2.

Phase 3

Eight missions were conducted during Phase 3 between June 14 and October 13, 1971. During this phase, color IR photography was collected every 14 days over all 210 segments and multispectral measurements were collected every 14 days over the 30 segments in the Intensive Study Area. Early in each 14-day period, ground observations of up to 12

corn fields in each segment were acquired. These data were processed and sent to the data reduction center at LARS. Fifteen segments of multispectral data and accompanying ground observations were sent to the data reduction center at WRL. The photographic and multispectral data were analyzed and results were reduced and reported to SRS in Washington and to the other participants in the Corn Blight Watch Experiment. The major data transfers for Phase 3 are shown in Figure 3.

During Phase 3 a new mission was started every other Monday, June 14, June 28, July 12, July 26, August 9, August 23, September 6, and September 20. Each mission was completed in 21 days and results were punched, checked, collated, and reported 23 days after the mission had began.

Only 14 days were scheduled for the collection of color IR and multispectral data. Data were collected initially over segments when cloud cover was 30% or less. If time and weather permitted, reflights were made when data on initial collection over segments resulted in more than 10% cloud cover. All such repeated flights were made when segments in question were expected to be covered by clouds 10% or less.

As in Phase 2, color IR photography (film type 2443) was collected at a scale of 1;120,000 over 36 flightlines. NASA/MSC identified the frame numbers to be analyzed and indicated the best frames when reflights were taken. NASA/MSC sent two transparencies and two positive contact prints of all color IR photography to the data reduction center at LARS.

The WRL aircraft collected multispectral data over the 30 segments in the Intensive Study Area. All data over the segments were checked at the data reduction center at LARS and immediately sent to the analysis center for processing.

When required by the analysis teams, low-altitude, large-scale photography was collected over a number of segments within the Intensive Study Area. These data were analyzed in conjunction with ground measurements to establish the exact condition of a number of fields. This information was used both to evaluate the performance of interpretation or of machine processed data results and to determine the source of any difficulty in data reduction.

Six to twelve corn fields in each of the 210 segments were designated by SRS to be visited by CES and data forms were distributed to each enumerator. Their biweekly reports were sent to the SRS state offices during the first week of the period, and, for the 30 flightlines in the Intensive Study Area, results were phoned to the Data Reduction Center at LARS and WRL. In each of 24 segments, up to 6 fields were designated to be visited by ASCS enumerators to provide test field information for data reduction results. These reports were also channeled to the SRS state offices where they were edited, coded, and punched onto data cards. At SRS in Washington, they were error checked copied onto digital data tapes, and listed on ground observation printouts. The

Table 1.

Principle Center of Data Acquisition and Processing

NASA/MSC	- NASA Manned Spacecraft Center, Houston, Texas
WRL/U. of Michigan	- Willow Run Laboratories, University of Michigan aircraft system.
ASCS County	- Agricultural Stabilization and Conservation Service of USDA County Offices
SRS/Washington	- Statistical Reporting Service of USDA, State Offices
SRS/State	- Statistical Reporting Service of USDA, State Offices
CES/State	- Cooperative Extension Service of the seven states
CES/County	- Cooperative Extension Service County Agency
DRC/WRL	- Data Reduction Center, Willow Run Laboratories, University of Michigan
DRC/LARS	- Data Reduction Center, Laboratory for Applications of Remote Sensing, Purdue University
USDA Information Center	- United States Department of Agriculture: Agricultural Information Center, Washington D. C.

biweekly data were delivered to the Data Reduction Center, and Ground Observation Summaries, described later, were distributed to the analysis teams by day 10 of each biweekly period, the same day as photographic data were available.

Data Analysis

The flow of remote sensor and ground collected data for the CBW was described above. Machine analysis of airborne multispectral scanner data was one part of the analysis effort. Other analysis efforts included photointerpretation of the high altitude color infrared photography and statistical analysis of the ground observations. Only the computer processing of the multispectral scanner data will be discussed here.

The multispectral scanner used in the CBW experiment was developed and is operated by the University of Michigan Willow Run Laboratories under contract to NASA. The scanner senses eleven bands in the wavelength region of .4 to 3 micrometers and one thermal infrared band from 9.7 to 11.3 micrometers. After sampling and digitation of the scanner radiometric measurements, the strip of terrain observed is represented by 222 samples across track and 176 samples per mile along track. At a 5000 feet altitude, the scan width is approximately two miles. A description of the scanner system is presented in reference 2.

Scanner data was recorded on analog tape, converted to digital form at LARS and stored on tape in 9 track 800 BPI format. An IBM system 360 Model 67 computer was used to perform all data processing functions. In order to relate ground truth data and aerial photography to aircraft scanner data a digital display system has been developed which recreates an image from the digital imagery stored on tape. The system includes a photocopy unit which enables hard copy output to be obtained. The image is presented on a 17 inch black and white screen in 16 gray levels by a raster of 528 lines with 768 points per line. A standard television refresh rate of 30 frames per second is maintained via a multihead disk buffer to offer a flicker free image. A lightpen is included for selection of points in the image. The system was developed by IBM FSD and is described in reference 3.

The first step in the analysis of the scanner data was to use the digital display to identify the location of the corn and non-corn fields to be studied in each segment. Figure 4 illustrates the use of the digital display for this purpose. An annotated aerial photograph was employed as a reference for each of the 30 scanner data test sites. Fields were located by the operator on the screen and the lightpen was used to mark the corners of the field. The field coordinates were punched on cards by the display software for use by the analysis programs. Test sites were nominally 1 mile wide and 8 miles long. Up to 200 agricultural fields were identified for each site. Field identification data was added to the coordinates on the field cards and the completed deck formed the basic analysis input along with the image data tapes.

The next step in the analysis procedure was to use a clustering procedure[4] on the data points in the fields. For this process an algorithm was used which subdivides a set of multidimensional samples into a given number of subsets in a way as to minimize the overall variance of the points in the subsets and to maximize the euclidian distance between the subsets. This algorithm is part of a remote sensor data analysis software system called LARSYS[5] which supplies statistical analysis, pattern classification, results display and other computational capability to users. A block diagram of the elements of the system is presented in Figure 5.

The LARSYS system is controlled by a monitor (see Figure 5) which provides a convenient interface between users and the various processors. The clustering processor reads "field" cards produced by the digital display system, reads the enclosed data points from tape, performs the clustering process and prints out a map of the results. The statistics processor computes first and second order multidimensional statistics, histograms, and spectral plots of data point sets identified by "field" cards. The feature selection processor determines the optimum feature subset to use for classification of samples drawn from populations described by the statistics produced by the statistics processor. These features are then used by the point classification or the sample classification processor to classify remote sensor data using the gaussian maximum liklehood ratio decision rule. The display processor produces classification maps and results tabulations for evaluation of the classifier performance. A results tape is written for archiving of the classification.

The results of the clustering were used to define training samples for subclasses of corn and other materials in

the scene. In the case of corn, the subclasses represented different blight levels, different crop conditions and variations due to angular dependancies across the flight line. In the case of other cover types, the subclasses represented wheat, soybeans, pasture, trees, etc. and variations of these classes. An example of clustering output in pictorial form is illustrated in Figure 6 for a block of agricultural materials including corn, soybeans and pasture.

After the training points were selected, to define the various classes of interest the feature selection processor of LARSYS was used to select the optimum set of channels for classification from the 12 available. Previous results of research had shown that four channels was a preferred number for agricultural materials of the type under study. The feature selection program often selected channels 5, (.58 - .60 micrometers); Ch. 6 (.58 - .65 micrometers); Ch. 7 (.61 - .70 micrometers); and Ch. 9 (1.0 - 1.4 micrometers). Variations in the selected set was observed during the period of the experiment however, and no one set of four was used for all classifications. Training statistics (means, variances and covariance matrices) were computed for each class using the LARSYS statistics processor and each test site was then classified. The classification results were tabulated for the total area classified into each blight class for each test segment.

Fifteen of the 30 test segments in the Intensive Study Area were analysed by LARS Purdue. At the University of Michigan Willow Run Laboratories, a similar analysis procedure was followed using analog techniques on the other 15 segments in the Intensive Study Area. Results for all 30 segments were reported both on a total segment basis and on a field-by-field basis with LARS and WRL using the same reporting forms. Each of the 30 segments were analyzed using these computer analysis proceedures for each mission which occured every fourteen days for a total of eight missions.

The correlation of segment average blight levels for two categories as estimated by computer classification in multispectral scanner data is shown in Figure 7. The data are number of acres in each class for each of the 30 segments in the intensive study area. Correlation coefficients were .94 and .92 for the two classes and the points lie close to the 1:1 line. As was true for the photointerpretive methods, attempts to classify the number of acres in each individual blight level were less accurate than for either two or three class groupings.

Corn Blight Record

A record of the information obtained for every field in each of the 210 segments was maintained on digital tape. The system designed for accomplishing this task and implemented for the Corn Blight Watch Experiment was called the Corn Blight Record and is shown in Figure 8. The initial interview data and biweekly field observations were merged with flight log information, multispectral analysis results, and photointerpretation results. The resulting tapes, one for the seven state area and one for the intensive study area, were the source of most of the listings and tabulations generated during the experiment.

Data analysis results in the form of remote sensing analysis tabulations were generated on day 23 of each period for the SRS in Washington. Expansion of results according to the sampling model for ground observations, photographic analysis, and multispectral data analysis were also generated for each period. Breakdowns of blight results for cytoplasm and for many other parameters were made. Yield calculations and other such studies were made. In addition, the results were analyzed using standard statistical techniques such as correlation, analysis of variance, and others.

Results Summarization and Dissemination

In Figure 9, the data flow for dissemination of the blight analysis results is summarized. For each biweekly period, color IR photographs were sent to the county enumerators for their particular segments, questions and training materials were sent with the prints, and results were returned for analysis and evaluation. The purpose of this aspect of the experiment was to acquaint the enumerators with small-scale photography in preparation for future technology.

Summarization of ground observations was performed by SRS within one week of data collection. Photographic and multispectral analysis results were sent to SRS within an average of two weeks after data was collected. These results were available to SRS for compiling blight reports to the USDA information center which in turn handled press releases to the news media.

Conclusions

Accurate estimates of the acerages of healthy and blighted corn in the Intensive Study Area were obtained from the machine analysis of the multispectral

scanner data. There was high correlation and agreement between ground estimates and machine analysis estimates. Analysis of multispectral scanner data gave a more accurate assessment of the blight situation than that provided by photointerpretation methods when compared with expanded ground observations. Corn was identified with a high degree of accuracy by both photointerpretive and machine analysis methods.

In conclusion, this near-operational test of remote sensing systems rapidly advanced our knowledge of their potential. In addition, it is expected that the data collected will continue to be useful in future research. Data over an agriculturally important area of the country have been collected through a growing season. More than 40,000 fields were included in the initial interview records. Ground observations were obtained for 1,600 corn fields visited biweekly. The results of photointerpretation for 16,000 corn fields were recorded every two weeks, and over 300 square miles of multispectral scanner data were analyzed eight times during the growing season. The procedures designed for handling the large amounts of data were successful. Where problems were encountered, adjustments were made to insure maximum results, available promptly and in a way that was consistant with the rapidly advancing state-of-the-art.

The 1971 Corn Blight Watch Experiment was planned by representatives of, and supported by, the experiment participants. The procedures summarized herein are a product of the total experimental plan; however, special appreciation goes to Mrs. S. K. Hunt and the Applications Programming group of LARS for their contributions in data management aspects of the experiment. Their work and that of other LARS staff to the experiment was supported by NASA under Grant NGL 15-005-112.

Portions of the material presented in this paper were presented at the NASA Fourth Annual Earth Resources Program Review, Houston, Texas, January 17 - 21, 1972. Further information on the CBW can be found in the proceedings of the Review to be published by NASA.

References

1. Bauer M.E., P.H. Swain, R.P. Mroczynski, P. E. Anuta, R. B. MacDonald, "Detection of Southern Corn Leaf Blight by Remote Sensing Techniques", Proceedings of the 7th International Symposium on Remote Sensing of Environment, University of Michigan; Ann Arbor, Michigan; May 17-21, 1971.

2. Fu, K. S., D. A. Landgrebe, T. L. Phillips, "Information Processing of Remotely Sensed Agricultural Data", Proc. IEEE, Vol. 57, pp. 639-653, April 1969.

3. IBM 4507 Digital Image Display System, IBM Federal Systems Division; Gaithersburg, Maryland; October 1970.

4. Wacker A. G., D. A. Landgrebe,"Boundaries in Multispectral Imagery by Clustering," 1970 IEEE Symposium on Adaptive Process, pp X14.1 - X14.8, December 1970.

5. "Remote Multispectral Sensing in Agriculture, "Purdue University Research Bulletin 873, Agricultural Experiment Station, Purdue University, Lafayette, Indiana, December 1970; also, Laboratory for Applications of Remote Sensing Annual Report, Volume 4, LARS, 1220 Potter Drive, West Lafayette, Indiana.

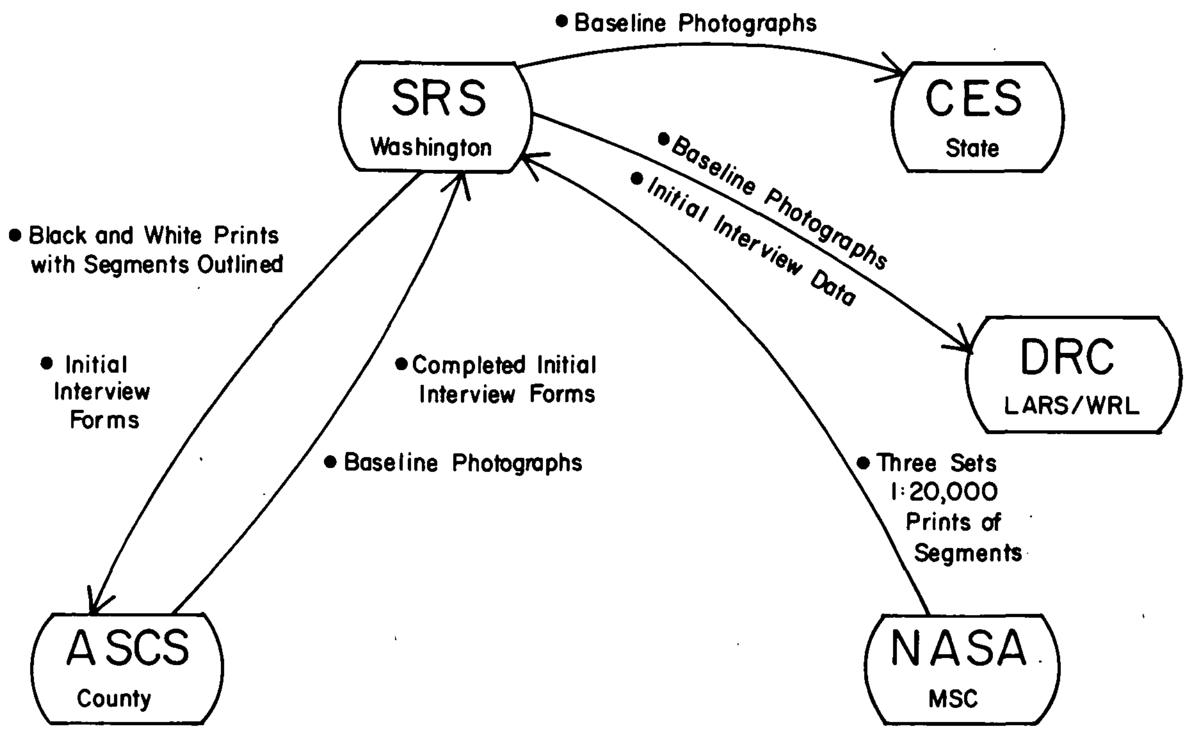

Figure 1. The data flow for obtaining baseline information is designed to provide baseline photographs to the Cooperative Extension Serivce of each state and baseline photographs and initial interview data to the Data Reduction Center by May 15, 1971. This aspect of the 1971 Corn Blight Watch Experiment was called Phase I, the first of three phases.

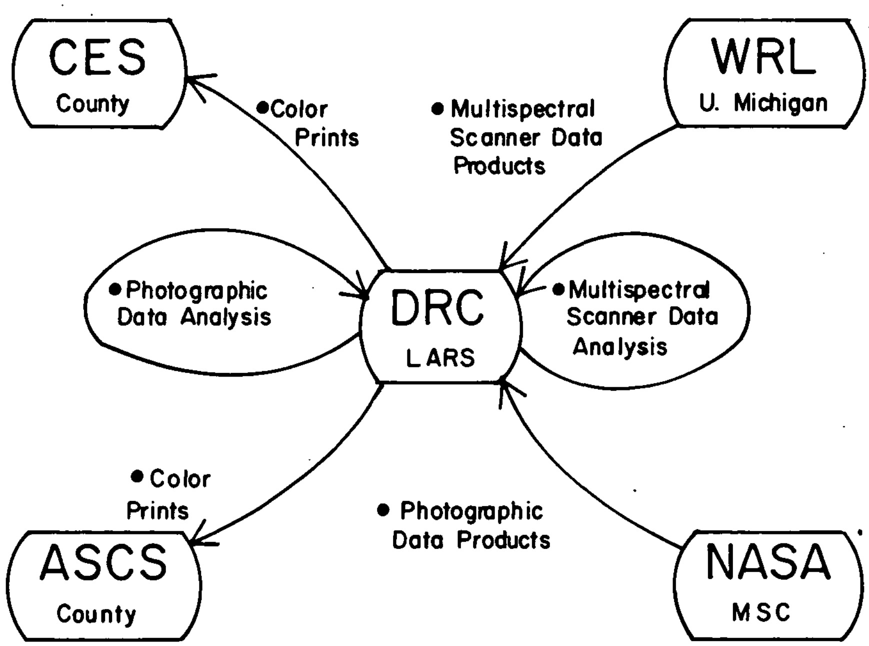

Figure 2. Phase II of three phases for the 1971 Corn Blight Watch Experiment was designed to collect data and perform analysis to determine soils background information for the corn fields. This data flow diagram shows how the objective was completed.

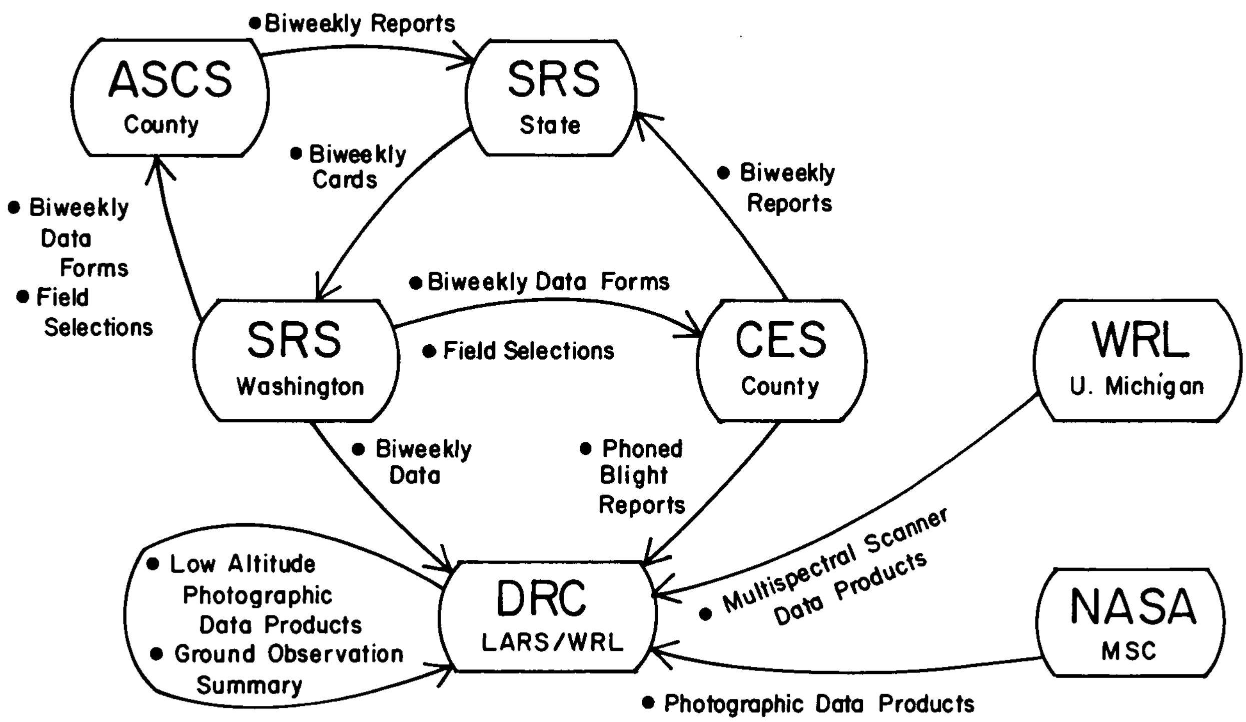

Figure 3. This data flow diagram for the last phase (Phase III) of the 1971 Corn Blight Watch Experiment shows the data acquisition for the experiment. The principal data products are biweekly field data, photographic data, and multispectral scanner data.

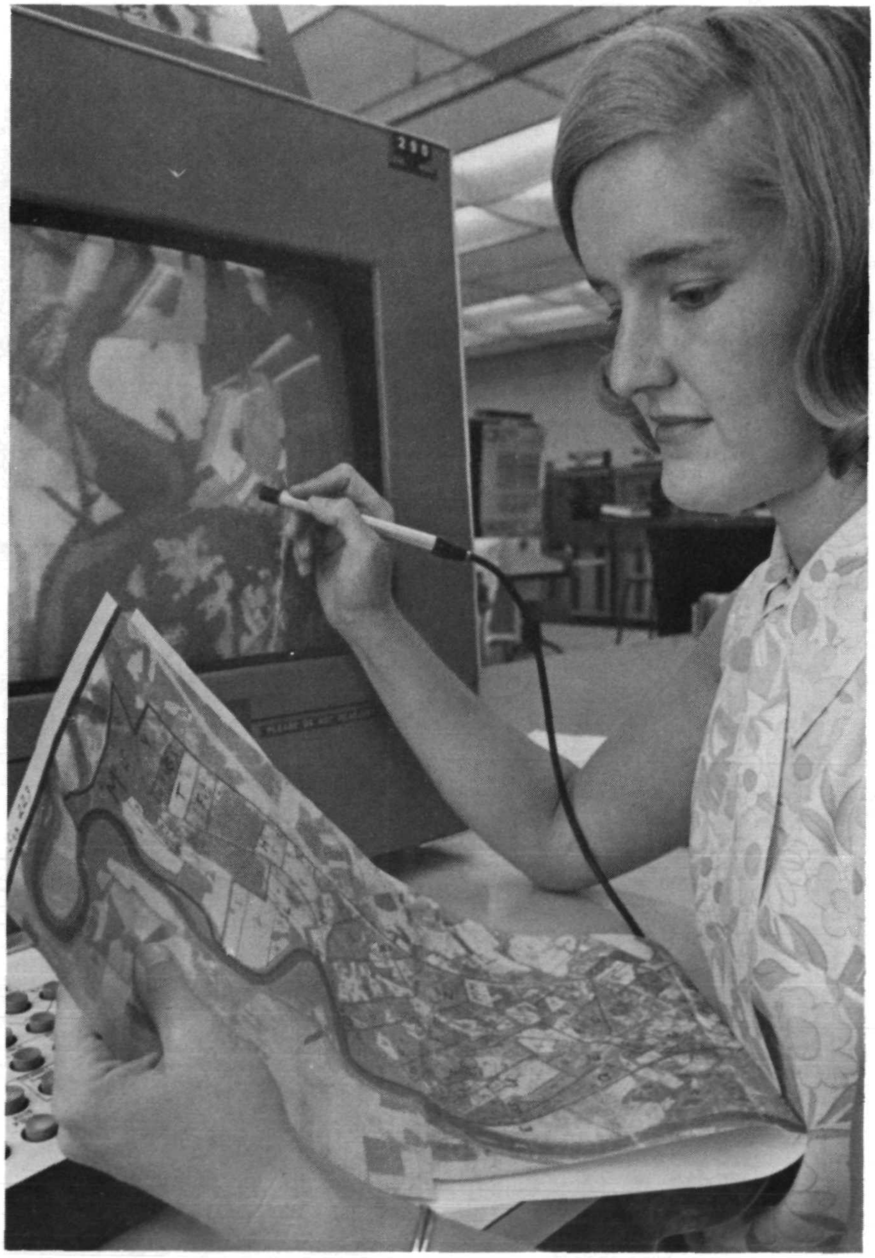

Figure 4. Digital images were used to locate training fields for classification of data from the multispectral scanner.

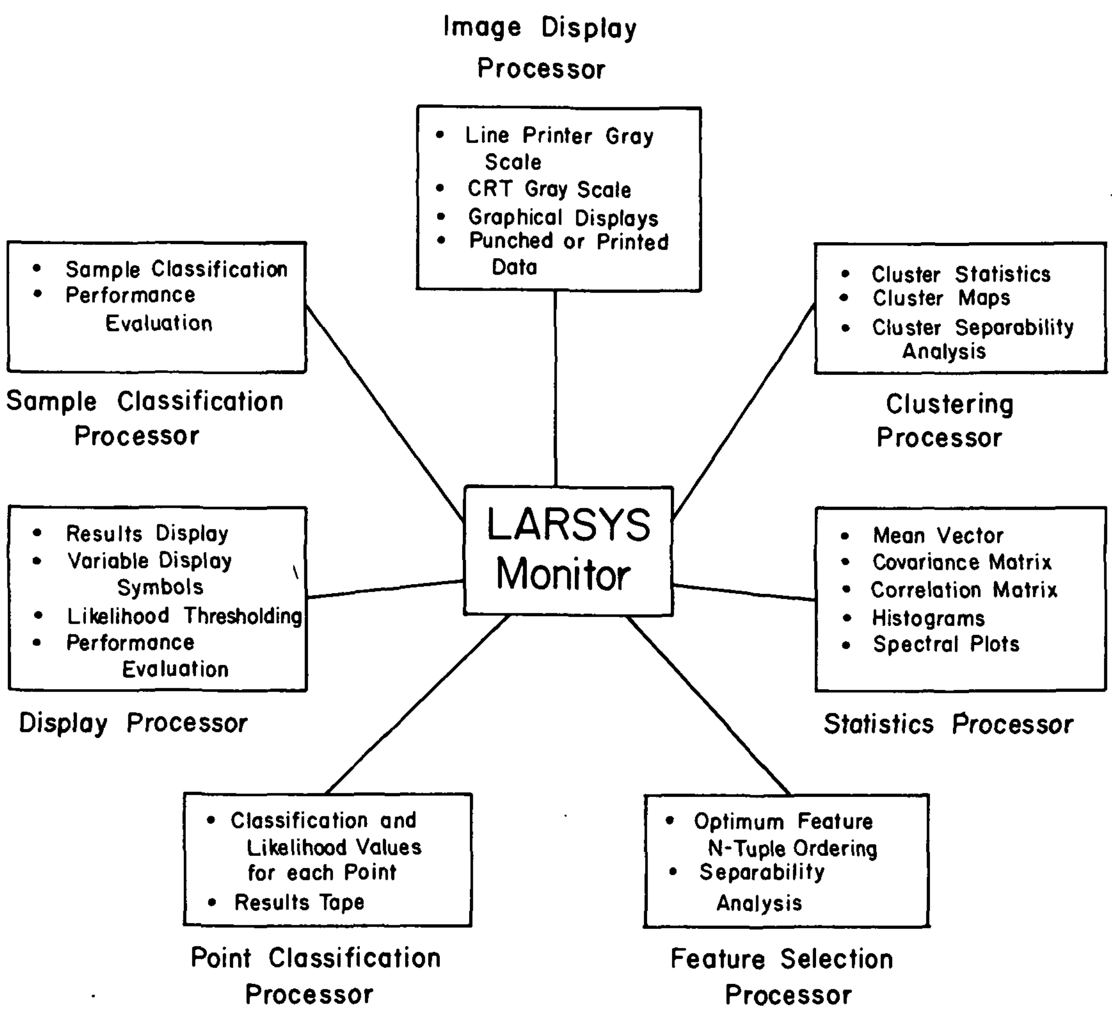

Figure 5. LARSYS: a software system for the analysis of multispectral remote sensing data.

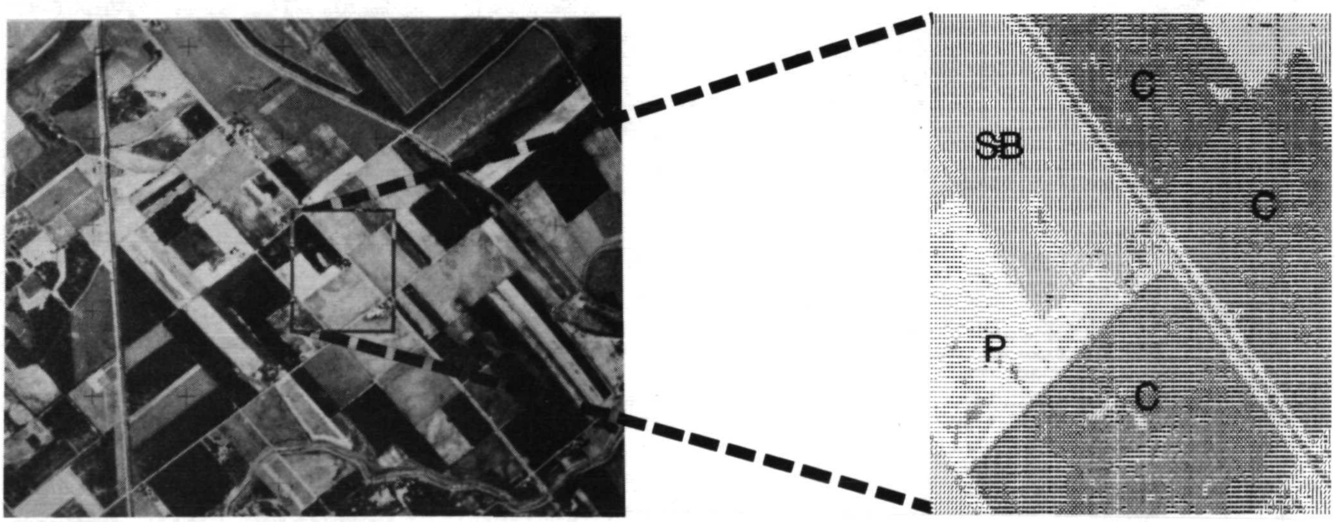

Figure 6. Example of Clustering Program Output. The area bordered in the digital display print on the left containing soybeans, corn, pasture, etc. was clustered and the symbol map on the right was printed out. The program accurately separated the crop and other cover types.

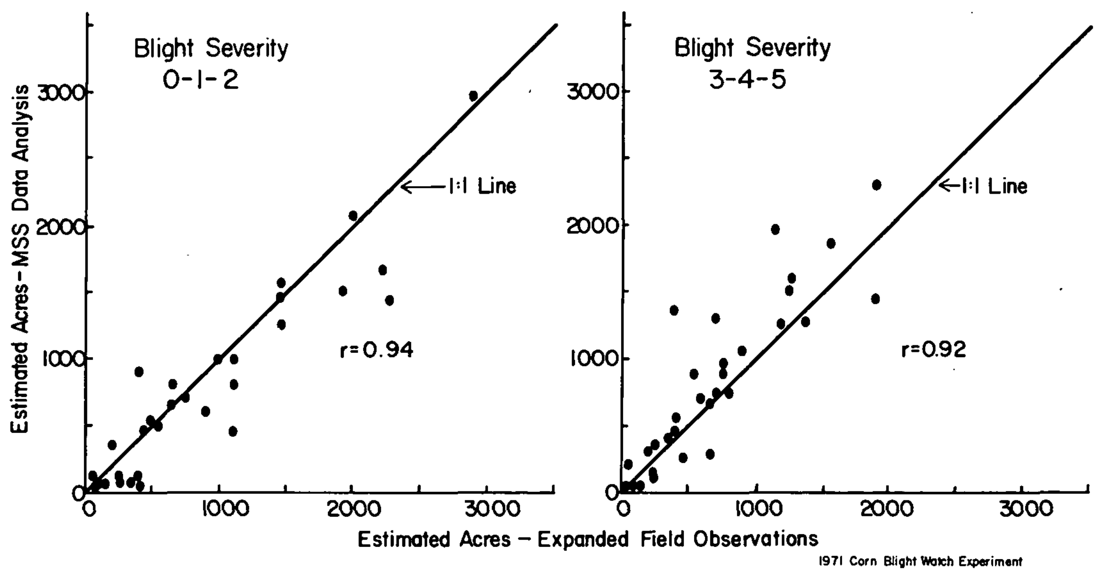

Figure 7. Correlation of estimates from multispectral scanner data and field observations of acreages of healthy (blight levels 0-1-2) and blighted (3-4-5) corn in the intensive study area for the period beginning August 9, 1971.

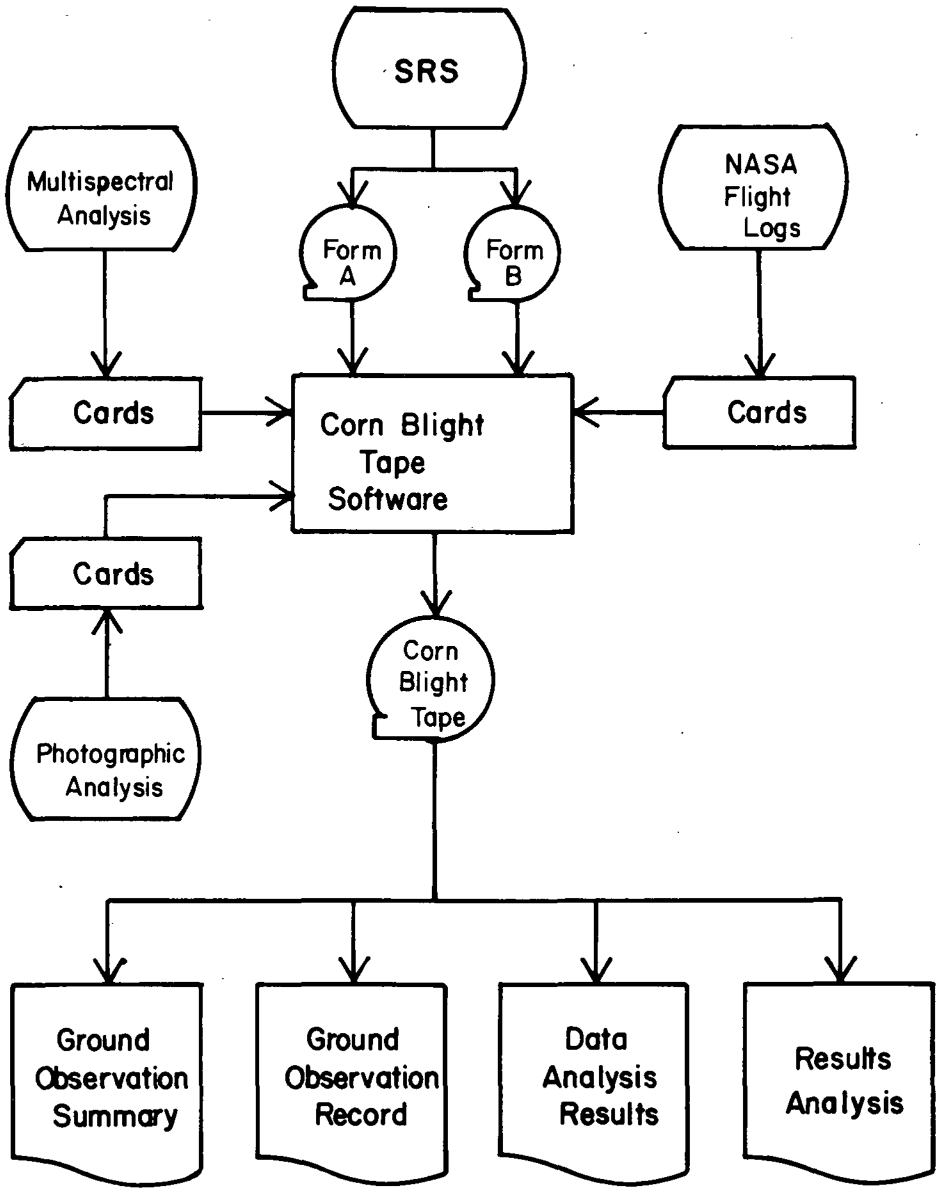

Figure 8. All ground data, flight logs, and analysis results were stored on magnetic tape. These merged data were used to report and analyze results for the 1971 Corn Blight Watch Experiment.

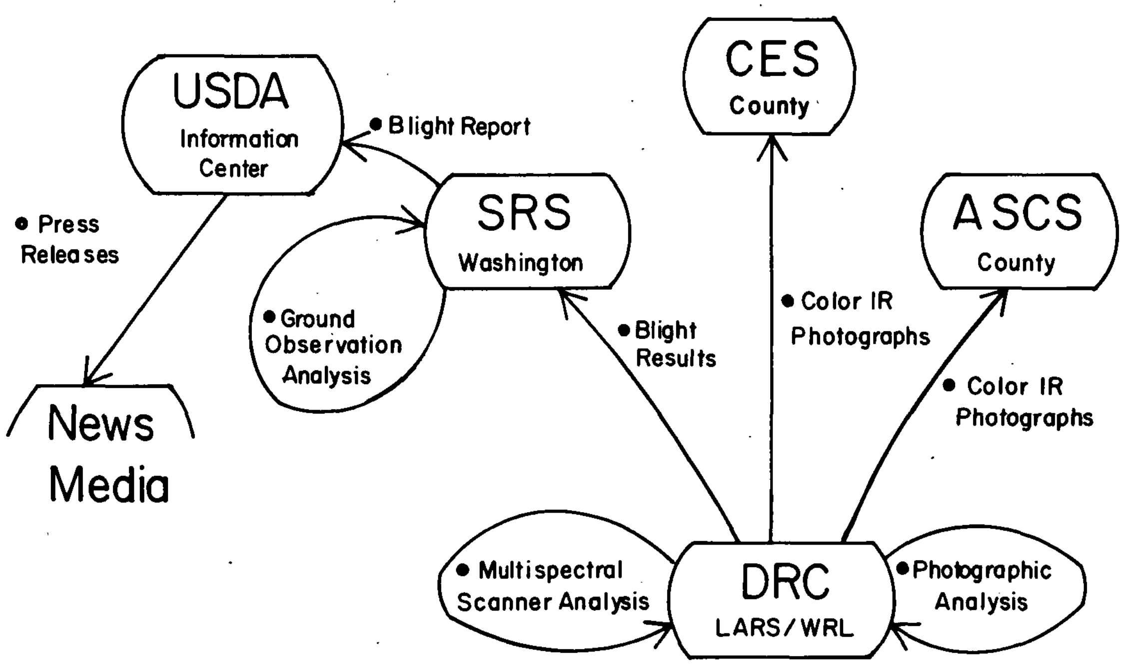

Figure 9. The data flow for results dissemination is diagrammed for Phase III of the Corn Blight Watch Experiment. This last phase of the experiment included the reporting of results to the public.